AWARD NUMBER: W81XWH-14-1-0097

TITLE: "Database for Parkinson Disease Mutations and Rare Variants"

PRINCIPAL INVESTIGATOR: JEFFERY M. VANCE

CONTRACTING ORGANIZATION: UNIVERSITY OF MIAMI
Coral Gables, FL 33146

REPORT DATE: September 2016

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
September 2016	FINAL	1 Jul 2014 – 30 Jun 2016

4. TITLE AND SUBTITLE
Database for Parkinson Disease Mutations and Rare Variants

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0097

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jeffery M. Vance, MD, PhD, Karen Nuytemans, PhD

E-Mail: jvance@med.miami.edu; knuytemans@med.miami.edu;

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Miami
1400 NW 10 Ave, Room 1007P
Miami, FL 33136-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT Massive parallel sequencing (MPS) allows for fast, high-throughput detection of rare variants, greatly increasing the research field's potential to study the 'common disease, rare variant' hypothesis in complex disorders. For Parkinson Disease (PD) specifically, the variant databases currently available are incomplete, don't assess impact and/or are not equipped to deal with massive parallel sequencing data. In this proposal, we set out to create a user-friendly database providing assessment of potential relevance of variants to PD through a ranking score and easy access to summary data of MPS datasets. We combine all information available on sequence variants identified in PD patients, from literature, publically available MPS datasets and MPS datasets from collaborators. Sequence variants are ranked in each of three evidence levels (genetic and functional evidence from literature and evidence from in-silico analyses performed in-house). Each variant will then be placed in one of six ranking score categories indicating impact to PD based on the strength of evidence from the three evidence-levels. Currently, variant positions, in-silico analyses data and ranking scores of the literature variants and two major MPS datasets (Miami Udall Center and the Progressive Parkinson Biomarker Initiative whole exome sequencing projects) have been uploaded to the flexible back-end structure of the database. Scripts for high-throughput analyses and data uploads to the database for future MPS datasets have been developed. All information is available through the Variant Database of Parkinson Disease (VarDoPa) website allowing for easy sharing of data and quick evaluation of identified sequence variants.

15. SUBJECT TERMS
Parkinson Disease; variant database; ranking score; sequence variants; online database; massive parallel sequencing; collaboration

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	24	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. **INTRODUCTION:** Massive parallel sequencing (MPS) allows for fast, high-throughput detection of rare variants, greatly increasing the research field's potential to study the 'common disease, rare variant' hypothesis in complex disorders. Due to the large influx of these data, however, information for the same variant is often spread across multiple sources and interpretation of the impact of the variant to disease is often difficult, especially for non-geneticists. For Parkinson Disease (PD) specifically, the number of variant databases currently available are incomplete, don't assess impact and/or are not equipped to deal with massive parallel sequencing data. In this proposal, we set out to combine all information available on sequence variants identified in PD patients from the literature, publically available MPS datasets and MPS datasets from collaborators. Sequence variants will be ranked in each of three evidence levels: 1) "Genetic evidence" (e.g. population frequency, family segregation) and 2) "functional evidence" extracted from literature and data repositories and 3) in-silico analyses for all variants to determine potential functional effect of the variants ("in-silico evidence") (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants). Each variant will then be placed in one of six categories indicating impact to Parkinson Disease based on the strength of evidence from the three evidence-levels (Table below). Summary data for each MPS dataset and/or for each sequence variant will be linked to the publication or to information on the original MPS laboratory. The inclusion of summary data only – not individual based data- allows for easier sharing of data and facilitate collaborative efforts. The use of the ranking category score will allow users to quickly evaluate the relevance of identified variants. All these data will be available through the Variant Database of Parkinson Disease (VarDoPa) website.

	Category 1	**Category 2**	**Category 3**
Genetic support	Highest	High	High
In-silico support	High	High or Intermediate	High or Intermediate
Functional support	Any	High	Intermediate or Low
	Category 4	**Category 5**	**Category 6**
Genetic support	Intermediate	Low	Low
In-silico support	Any	High	Intermediate or Low
Functional support	Any	Intermediate or Low	Low

2. **KEYWORDS:** Parkinson Disease; variant database; ranking score; sequence variants; online database; massive parallel sequencing; collaboration;

3. **ACCOMPLISHMENTS:**

Major goals of the project:

- Construct database structure for sequencing data of Parkinson Disease datasets
- Extract genetic information from literature for known and candidate PD genes

- Obtain in silico annotation data for in-house and outside MPS datasets as well as variants reported in literature
- Develop ranking score system
- Develop website supporting the database

Accomplished under these goals:

- Construct database structure for sequencing data of Parkinson Disease datasets

The back-end table structure for the database has been constructed and can accommodate multiple MPS datasets. We do not include individual level data, but require summary data per dataset (i.e. x carriers out of y individuals). Practically, this allowed for a less complex back-end database structure. Tables are organized based on dependency on genome wide position only or effect of variant within the gene context (March 2015).
We have developed scripts to *extract information out of the annotation files and correctly align variants with their respective annotations from different algorithms to populate the back-end tables (June 2015), *calculate frequencies across datasets and *identify novel variants for annotation in to-be-added datasets (December 2015). The back-end tables are flexible and can be adjusted to include more variants, more annotations and more summary data across different datasets. Additional scripts were developed (December 2015) to allow for easy automated updating of the back-end tables as data on new variants becomes available.

- Extract genetic information from literature for known and candidate PD genes

Literature searches included extracting information on frequency observed in PD patients, segregation within PD families and functional analyses performed to test the effect of the observed variant on the encoded protein's function. (a) We obtained all the literature information available through the Belgian website PDmutDB (http://www.molgen.vib-ua.be/PDMutDB/). (June 2015) (b) Additional data extraction was needed for those reports included in the PDmutDB database, as PDmutDB does not report on number of individuals screened (i.e. x carriers out of y individuals) (December 2015). (c) We completed additional literature searches for the known PD genes (*SNCA*, *PARK2*, *PINK1*, *LRRK2*, *ATP2A13*, *FBXO7*, *PLA2G6*, *HTRA2*, *EIF4G1*, *VPS35*). (June 2015) (d) So far, we have completed literature searches for up to 26 candidate genes (June 2016). These searches have been time consuming and will continue past the end date of this award. As more data becomes available, previous searches will be repeated to include all recent information.

- Obtain in silico annotation data for in-house and outside MPS datasets as well as variants reported in literature

(a) In-silico analyses include standard annotation (e.g. function within gene, conservation and frequency in the general population) through SeattleSeq (http://gvs.gs.washington.edu/SeattleSeqAnnotation/) and annotation specific for each genre of variants. Coding variants are assessed for their impact on the protein function (amino acid change; PolyPhen2[1], SIFT[2] and Provean[3]/ presence in functional domains; Variant Effect Predictor[4]) and gene splicing (MutPredSplice[5] and ESEfinder[6]). Splice variants are tested for their ability to change gene splicing using several prediction programs (MutPredSplice[5], Human Splice Finder[7], MaxEntScan[8], NNSplice[9] and SpliceView[10]). Noncoding variants (intergenic, intronic and untranslated region variants) are assessed for

regulatory potential through programs investigating specific binding sites (miRNA; TargetScan[11] and MiRBase[12], and transcription factors; ENCODE) and algorithms estimating potential of the region to affect gene expression by interpreting the publically available data (e.g. ENCODE, gene expression and binding site data; CADD[13], RegulomeDB[14], Funseq2[15] and GWAVA[16]).
(b) A standard operating protocol for in-silico annotation has been developed (June 2015). This includes scripts to extract the genres of variants in separate files to allow for more streamlined additional analyses through variant specific input. Additional programming was needed for the MPS data, as we discovered that many annotation programs are not equipped to deal with the high amount of data coming from MPS datasets. To address this problem, we have developed protocols to divide the datasets upfront to overcome this obstacle. In addition, the available splice prediction algorithms were developed before onset of MPS data and none present itself as a 'golden standard'. For this reason we have opted to include a summary analysis across several prediction programs.
(c) Annotation of the literature data from PDmutDB (December 2014) and two MPS datasets (in-house 'UMiami WES' (June 2015) and Parkinson's Progression Markers Initiative 'PPMI WES' (July 2016)) is completed and uploaded in the back-end database. We had moved the inclusion of the in-house MPS data ahead of the initial plan, as we realized it would be advantageous in working through the development of the MPS part of the database.

- Develop ranking score system

Back-end algorithms were fully developed to determine each variant's level of evidence in each of the three categories ('genetic', 'in silico' and 'functional' evidence) and overall rank score for impact in PD (March 2016). Genetic evidence level is based on frequency in general population databases, available information on co-segregation with disease in families and number of reports in affecteds versus controls. Functional evidence level is determined by assessing number of functional assays reported for the variant and the consistency of the effect observed. For each variant type (nonsense, missense, splice, synonymous, UTR and intronic/intergenic), in silico evidence level is obtained through assessment of results of algorithm described in 'in silico annotation data' section above. These three evidence levels are then combined to define an impact rank for the variant in PD. Ranking score is now available for all literature data and both MPS datasets (July 2016).

- Develop website supporting the database

(a) The website is available through the John P. Hussman Institute for Human Genomics' website; http://www.hihg.med.miami.edu/vardopa (Sept 2015). The front-end of the VarDoPa website has been designed to allow for personalized lists of annotation to be included on the screen. Each annotation column has the option for additional filtering, and all data is available for download to the user's personal computer through the export function. Initial testing to identify small errors in the usage of the website and its functionalities has been completed.
(b) Extra pages supporting the database have been added to the website (Dec 2015). Each dataset will be linked to the information on the contributing research group and the pipeline used to generate the variant dataset (http://www.hihg.med.miami.edu/vardopa/dataset-

information). Specific information on the annotation algorithms (including which ones, versions used and parameters used) is available and will be updated as changes are made or algorithms are added (http://www.hihg.med.miami.edu/vardopa/annotation-information). Additional information on Parkinson Disease in general and the rationale behind VarDoPa can be found in the 'About' pages (http://www.hihg.med.miami.edu/vardopa/about-parkinson-disease/ http://www.hihg.med.miami.edu/vardopa/about-vardopa).

Opportunities for training and professional development: Nothing to report.

Dissemination of results to communities of interest: Preparation of manuscript is ongoing.

Plans for next reporting period: Nothing to report; this is the final report.

4. IMPACT:

Impact on the development of the principal discipline of the project:

Within the last few years, it has become clear that it is important to understand what variants are potentially contributing to PD, both short term for researcher to focus their studies and long term for the medical field as it's moving towards precision medicine. As VarDoPa catalogs all rare variants reported in PD sequencing research, mostly encompassing rare variants, it extends beyond existing databases such as PDGENE (common variants) and Human Gene Mutation Database and Leiden Open Variant Database (literature reports only).

Sharing data across research groups is often difficult however due to informed consent or IRB restrictions. VarDoPa does not include individual level data, but uses summary data across datasets. This will allow researchers to share their data in a much more straightforward fashion. Alternatively, variants presented on VarDoPa will be linked to the contact information of the original laboratory in which the variant was identified, allowing for easier collaborative efforts.

The database has two major goals within the Parkinson disease field; (1) easier opportunity for collaboration between researchers as no individual data needs to be shared upfront for inclusion in the database and (2) more straightforward interpretation of the potential impact of each variant in PD development through summarization of available annotations.

Impact on other disciplines: Nothing to report.
Impact on technology transfer: Nothing to report.
Impact on society beyond science and technology: Nothing to report.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change:

Throughout the grant period we did not make any large changes that didn't fall within the scope of the original proposal. Between grant proposal and grant award, it became clear that a large amount of MPS data would become available earlier in the grant's timeline. The concerns presented in the original proposal in terms of interpretation of the variant data and sharing of data are also valid for MPS data. Therefore, we expanded the scope of the proposed database from literature sequencing data to all sequencing data, including

larger MPS datasets. This changed the order of our objectives across the two years but no goals were changed.

Actual or anticipated problems or delays and actions or plans to resolve them:
Two distinct delays were accrued. In silico annotation of large datasets was more time intensive than anticipated as many annotation programs were developed before the onset of MPS and thus do not allow for high-throughput screening. Secondly, literature searches required more time than originally anticipated; especially older literature before naming conventions became standard.

Changes that had a significant impact on expenditures: Nothing to report.
Significant changes in use or care of human subjects, vertebrate animals, biohazards and/or select agents: Nothing to report.

6. PRODUCTS:

Publications, conference papers, and presentations:

Conference abstracts:

- Nuytemans, K., Wang, L., Beecham, G.W., Vance, J.M.: "Database of evidence-based ranked Parkinson Disease variants". NINDS 16th Annual Udall Center Directors' Meeting, Rockville, MA, October 30-31, 2014.
- Nuytemans, K., Wang, L., Beecham, G.W., Van Broeckhoven C, Vance, J.M.: "Parkinson Disease Variant Database (PDVD) with multiple evidence levels to rank the significance of sequence variants". 12th International Conference on Alzheimer's and Parkinson's Diseases (AD/PD), Nice, France, March 18-22, 2015.
- Vance, J., Wang, L., Beecham, G., Van Broeckhoven, C., Nuytemans, K.: "Database of evidence-based ranked Parkinson Disease variants". American Academy of Neurology 67th Annual Meeting, Washington, DC, April 18-25, 2015.
- Nuytemans, K., Wang, L., Beecham, G.W., Van Broeckhoven C, Vance, J.M.: "Parkinson Disease Variant Database". MDS 19th International Congress of Parkinson's Disease and Movement Disorders, San Diego, CA, June 14-18, 2015.
- Vance, J.M., John-Williams, K., Ali, A., Mehta, A., Wang, L., Beecham, G.W., Van Broeckhoven, C., Nuytemans K.: "Significance ranking of sequence variants in Parkinson Disease variant database". American Society of Human Genetics, Baltimore, MD, October 6-10, 2015.
- Nuytemans, K., John-Williams, K., Ali, A., Mehta, A., Wang, L., Beecham, G.W., Van Broeckhoven, C., Vance, J.M.: "Significance ranking of sequence variants in Parkinson Disease variant database". NINDS 17th Annual Udall Center Directors' Meeting, Washington, DC, October 27, 2015. (no abstract needed)

Presentations:

- Nuytemans, K.: "Insights into the complex nature of Parkinson Disease through next generation sequencing". Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, December 1, 2014
- Nuytemans, K.: "Identifying variants in big datasets and assessing their involvement in Parkinson Disease". Neurology Grand Rounds, Mayo Clinic Jacksonville, Jacksonville, FL, March 16, 2015

Website(s) or other internet site(s):
Creation of variant database: http://www.hihg.med.miami.edu/vardopa

Technologies or techniques: Nothing to report.
Inventions, patent applications, and/or licenses: Nothing to report.
Other products: Nothing to report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS:

Individuals that worked on the project:

Name:	*Jeffery M. Vance*
Project Role:	*PI*
Researcher Identifier (e.g. ORCID ID):	*eRA Commons: jvance*
Nearest person month worked:	*2*
Contribution to Project:	*Dr. Vance supervised the annotation and development of the backend structure and frontend website.*
Funding Support:	-

Name:	*Karen Nuytemans*
Project Role:	*Co-I*
Researcher Identifier (e.g. ORCID ID):	*eRA Commons: knuytemans*
Nearest person month worked:	*5*
Contribution to Project:	*Dr. Nuytemans has supervised the annotation, the implementation of the data into the backend structures and the development of frontend website. She has been created scripts and developed the SOP for the annotation.*
Funding Support:	-

Name:	*Liyong Wang*
Project Role:	*Co-I*
Researcher Identifier (e.g. ORCID ID):	*eRA commons: lwang55*
Nearest person month worked:	*3*

Contribution to Project:	*Dr. Wang has been involved in the evaluation of the splice algorithms and evaluation of the frontend website.*
Funding Support:	-

Name:	*Gary Beecham*
Project Role:	*Co-I*
Researcher Identifier (e.g. ORCID ID):	*eRA commons: gbeecham*
Nearest person month worked:	*1*
Contribution to Project:	*Dr. Beecham supervised the creation of the backend structure and frontend website as head of the Division of Research Informatics in the Center for Genetic Epidemiology and Statistical Genetics.*
Funding Support:	-

Name:	*Vanessa Inchausti*
Project Role:	*Research associate*
Researcher Identifier (e.g. ORCID ID):	-
Nearest person month worked:	*3*
Contribution to Project:	*Ms. Inchausti contributed to the project through annotation of the PDmutDB variants and in-house UMiami WES dataset, and initial literature searches.*
Funding Support:	-

Name:	*Elise Bendik*
Project Role:	*Research associate*
Researcher Identifier (e.g. ORCID ID):	-
Nearest person month worked:	*1*
Contribution to Project:	*Ms. Bendik contributed to the project through annotation of the in-house UMiami WES dataset.*
Funding Support:	-

Name:	*Krista John-Williams*

Project Role:	*Research associate*
Researcher Identifier (e.g. ORCID ID):	-
Nearest person month worked:	*9*
Contribution to Project:	*Ms. John-Williams completed the annotation for the in-house UMiami WES and PPMI WES dataset.*
Funding Support:	-

Name:	*Aleena Ali*
Project Role:	*Research associate*
Researcher Identifier (e.g. ORCID ID):	-
Nearest person month worked:	*9*
Contribution to Project:	*Ms. Ali contributed mostly to the literature searches of the known PD genes and candidate genes.*
Funding Support:	-

Change in the active other support of the PD/PI(s) or senior/key personnel since last reporting period: Nothing to report.
Other organizations involved: Nothing to report.

8. SPECIAL REPORTING REQUIREMENTS:
Collaborative awards: Nothing to report.
Quad charts: Attached.

9. APPENDICES:
- Quad chart
- Abstracts
- Presentation announcements

Database for Parkinson Disease Mutations and Rare Variants

Proposal Log #13298004
W81XWH-BAA-13-1

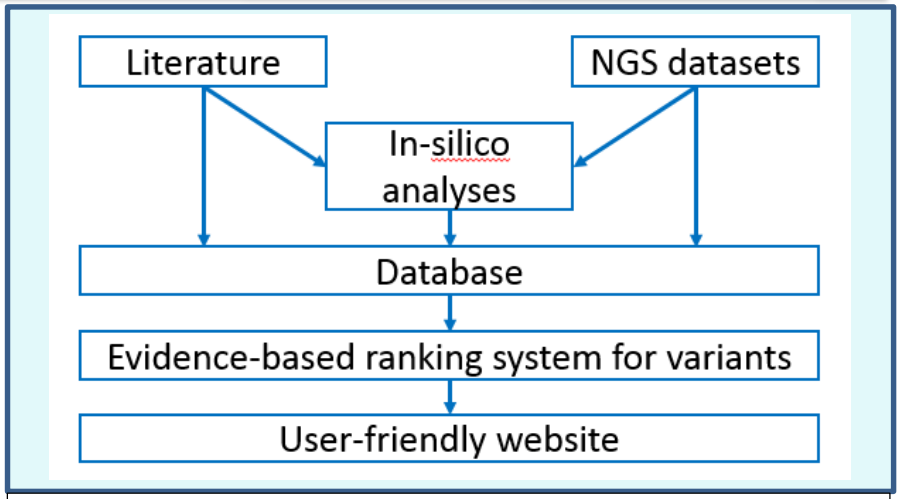
PI: Jeffery Vance **Org:** University of Miami **Award Amount:** $467,804.00

Study/Product Aim(s)

- Creation of web-based database of specific genomic variants in reported PD genes and new candidate genes
- Develop a ranking system for specific genomic variants using the basic concepts of evidence based medicine, literature and in-house in-silico analyses.
- Publicize website through publications and meetings.

Approach

We will go through PD literature and extract all genetic information on variants possibly involved with PD development ("genetic evidence" and "functional evidence"). In addition, we will include variant data from large NGS datasets. For all variants, "in silico" analysis will be performed to evaluate functional impact of the variant. All variants will be ranked based on the evidence available in the three categories mentioned above.



Ranking score and supporting data on literature variants and two large sequencing datasets is available through the VarDoPa website. Manuscript preparation is ongoing.

Timeline and Cost

Activities CY	14	15	16	17
Construct Database structure				
Populate database with literature and in silico data for causal/candidate PD genes				
Populate database with variants and in silico data for NGS datasets				
Develop/publish user friendly website				
Estimated Budget ($K)	**$133,590**	**$233,903**	**$100,311**	**$000**

Updated: 06/30/16

Goals/Milestones

CY14 Goals – Construct database structure/Populate database
- ☑ All information for known causal PD genes was extracted
- ☑ Development of website supporting the database was started
- ☑ Obtain in silico annotation data for causal PD genes
- ☑ Finalize database structure

CY15 Goals – Develop user friendly website/Populate database
- ☑ Populate database with variants and in silico data for in-house NGS datasets
- ☑ Develop website
- ☑ Populate database with variants and in silico data for large NGS datasets
- ⊡ Populate database with literature and in silico data for candidate PD genes
- ☑ Develop/Evaluate ranking system

CY16 Goal – Publish web-based database
- ⊡ Publish website and accompanying report

Comments/Challenges/Issues/Concerns

Ranking score and supporting data on literature variants and two large sequencing datasets is available through the VarDoPa website. Manuscript preparation is ongoing.

Budget Expenditure to Date

Projected Expenditure:
Actual Expenditure:

ABSTRACT NUYTEMANS UDALL 2014

"Database of evidence-based ranked Parkinson Disease variants"

Authors: Nuytemans, K.[1], Wang, L.[1-2], Beecham, G.W.[1-2], Vance, J.M.[1-2]

1) University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics; 2) University of Miami, Miller School of Medicine, Dr. John T. Macdonald Foundation Department of Human Genetics, Miami, FL 33136, USA;

At present, 15 major risk or causative loci have been identified for PD, though interpretation for each variant in these known genes can be difficult as many novel, rare or deleterious variants in these known genes as well as new genes do not lead to disease. The introduction of next generation sequencing has significantly increased the rate these variants are reported, changing the focus of research in PD from gene-based information to the single basepair variant. These factors make it increasingly difficult for the average PD researcher to identify those useful and important variants in the high throughput of NGS. The need for a new kind of genetic database, capable of ranking and handling NGS data and/or allowing for collaborations without sharing raw sequencing data was recently noted by the Udall PD research directors.

Current mutation databases are incomplete and/or do not do not tackle actual significance of single variants in the disease process. Further, most are not built to accommodate NGS data. The proposed work would fill this gap by creating a ranking scale for each variant based on current documentation and additional in-silico analyses and make recommendations about the strength of evidence for their significance to PD. Inclusion of summary NGS data on a research group level will allow for easy collaborative efforts.

We will gather evidence on three levels for each variant ("genetic", "functional" and "in silico"). We will go through literature and extract all genetic information on variants possibly involved with PD development including population frequency, family segregation, etc ("genetic evidence") and functional data, where available ("functional evidence"). We acknowledge that these data will likely not be available for NGS variants. We will perform in-silico analyses for all variants (reported or NGS) to determine the potential functional effect of the variants ("in-silico evidence") specific to their basic annotation (e.g. PolyPhen2 for nonsynonymous, RegulomeDB for UTR, GWAVA for non-coding variants,..). With these three-level data, we will develop a ranking system of 6 major categories, based on differential rankings within each of the 3 evidence classifications to provide a well-founded estimate of functional impact of each variant on PD development. All evidence will be available to the user through the website by the simple inclusion (at the user's discretion) of columns with the information-of-interest.

Summary data for each NGS dataset will be represented in its own separate column and will be linked to NGS pipeline and PI contact information to facilitate collaborative efforts.

The creation of this user-friendly mutation database for Parkinson Disease will allow the primary user to quickly identify variants as a possible causal or risk factor with variable strengths and to set up potential collaborations. In a grander scheme the database allows for integration of larger datasets (without the need for raw data sharing) and will improve our overall

understanding of the genetic basis for PD (both priority recommendations for basic research by NINDS PD2014).

ABSTRACT VANCE DOD AD/PD NICE, FRANCE 2015

"PARKINSON DISEASE VARIANT DATABASE (PDVD) WITH MULTIPLE EVIDENCE LEVELS TO RANK THE SIGNIFICANCE OF SEQUENCE VARIANTS"

Authors: Nuytemans, K.[1], Wang, L.[1], Beecham, G.W.[1], Van Broeckhoven C[2], Vance, J.M.[1]

1) University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics; 2) Department of Molecular Genetics, VIB and Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

Objective: Next generation sequencing (NGS) has significantly increased the rate of rare sequence variants (SV) reported in PD patients. But interpretation of the functional significance of SV can be difficult. Current databases do not address this question and/or do not accommodate NGS data. The PDVD is designed to address these needs.

Methods: SV included will come from the literature and existing NGS databases. SV are ranked in each of three evidence levels: 1) "Genetic evidence" (e.g. population frequency, family segregation) and 2) "functional evidence" extracted from literature and data repositories and 3) in-silico analyses for all variants to determine potential functional effect of the variants ("in-silico evidence") (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants).

Results: Each variant is placed in one of six categories, based on the strength of evidence from the three evidence-levels, and this data will be available through a website. Summary data for each dataset and/or for each SV will be linked to the NGS laboratory or publication, to facilitate collaborative efforts.

Conclusions: The PDVD will allow users to quickly evaluate variants and initiate collaborations. The summary data format avoids individual identifiers to allow rapid availability of NGS data to the PD research community. The database meets recommendations by NINDS PD2014. Funded by the US Department of Defense.

Table 1. Rankings for SV

	Category 1	Category 2	Category 3
Genetic support	Highest	High	High
In-silico support	High	High or Intermediate	High or Intermediate
Functional support	Any	High	Intermediate or Low
	Category 4	**Category 5**	**Category 6**
Genetic support	Intermediate	Low	Low
In-silico support	Any	High	Intermediate or Low
Functional support	Any	Intermediate or Low	Low

Database of evidence-based ranked Parkinson Disease variants

Authors: Jeffery Vance, Liyong Wang, Gary Beecham, Christine Van Broeckhoven, Karen Nuytemans

University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics

Objective: To create a web database that can provide interpretation of the significance of single variant changes (SV) in Parkinson Disease (PD).

Background: Evaluation of genetic variants in PD can be difficult, as even rare or deleterious variants may not lead to disease. The introduction of next generation sequencing (NGS) has significantly increased the rate these variants are reported, changing the focus of research in PD from gene-based to SV. Thus, there is a need for a new kind of genetic database, capable of ranking and handling NGS data. The proposed work would fill this gap by creating a ranking scale for each variant based on current documentation and additional in-silico analyses and make recommendations about the strength of evidence for their significance to PD.

Methods: SV included will come from the literature and existing NGS databases. SV are ranked in each of three evidence levels: 1) "Genetic evidence" (e.g. population frequency, family segregation) and 2) "functional evidence" extracted from literature and data repositories and 3) in-silico analyses for all variants to determine potential functional effect of the variants ("in-silico evidence") (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants).

Results: Each variant is placed in one of six categories, based on the strength of evidence from the three evidence-levels, and this data will be available through a website. Summary data for each dataset and/or for each SV will be linked to the NGS laboratory or publication, to facilitate collaborative efforts.

Conclusions: The database will allow users to quickly evaluate variants and initiate collaborations. The summary data format avoids individual identifiers thus allowing rapid availability of NGS data to the PD research community. The database meets recommendations by NINDS PD2014.

Study was supported by a grant from the Department of Defense.

ABSTRACT VANCE DOD MDS SAN DIEGO, US 2015

Authors: Nuytemans, K.[1], Wang, L.[1], Beecham, G.W.[1], Van Broeckhoven C[2], Vance, J.M.[1]

1) University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics; 2) Department of Molecular Genetics, VIB and Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

SIGNIFICANCE RANKING OF SEQUENCE VARIANTS IN PARKINSON DISEASE VARIANT DATABASE

Objective: Create variant database with ranking of significance per variant to Parkinson Disease (PD).

Background: Next generation sequencing (NGS) has significantly increased the rate of rare sequence variants (SVs) reported in PD patients. But interpretation of the functional significance of SVs can be difficult due to time or knowledge restrained to sift through scattered information. Current databases do not address actual contribution of variant to PD pathogenesis and/or do not accommodate NGS data. The PD variant database is designed to address these needs.

Methods: SVs included will be extracted from the literature (known and candidate genes) and existing NGS databases. SVs are ranked in each of three evidence levels: 1) "Genetic evidence" (e.g. population frequency, family segregation) and 2) "functional evidence" extracted from literature and data repositories (e.g. binding specificity assays, expression analyses) and 3) in-silico analyses for all variants to determine potential functional effect of the variants ("in-silico evidence") (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants).

Results: Each variant is placed in one of six ranks, based on the strength of evidence from the three evidence-levels [Table 1]. The data supporting the rank will also be available to the user through the website (proposed launch mid2015) and detailed description of the ranks will be provided to simplify interpretation of the ranks. Summary (quality control) data for each dataset and summary frequency data for each SV will be linked to the NGS laboratory's info or publication, to facilitate collaborative efforts.

Conclusions: The PD variant database will allow users to quickly evaluate variants when interpreting sequence data, leading to a more focused follow-up in a research or clinical setting. The summary data format avoids the presence of individual identifiers, facilitating rapid availability of NGS data to the PD research community and thus potential collaborative efforts.
Funded by the US Department of Defense.

Table 1. Rankings for SV

	Rank 1	Rank 2	Rank 3
Genetic support	Highest	High	High

In-silico support	High	High or Intermediate	High or Intermediate
Functional support	Any	High	Intermediate or Low
	Rank 4	**Rank 5**	**Rank 6**
Genetic support	Intermediate	Low	Low
In-silico support	Any	High	Intermediate or Low
Functional support	Any	Intermediate or Low	Low

ABSTRACT VANCE DOD ASHG BALTIMORE, US 2015

Authors: **J.M. Vance[1,2,3], K. John-Williams[1,2], A. Ali[1,2], A. Mehta[1,2], L. Wang[1,2,3], G.W. Beecham[1,2,3], C. Van Broeckhoven[4,5], K. Nuytemans[1,2]**

1) University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics, Miami Morris K. Udall Parkinson Disease Research Center of Excellence (UPDRCE); 2) Department of Molecular Genetics, VIB and Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

Significance ranking of sequence variants in Parkinson Disease variant database

Next generation sequencing (NGS) has significantly increased the rate of rare sequence variants (RVs) reported in PD patients. Interpretation of the functional significance of RVs in PD development can be difficult however, due to the vast amounts of data and disparate data sources. Current available databases do not address the contribution of variants to PD pathogenesis and/or do not accommodate NGS data. We designed the PD variant database to address these needs.

Variants are extracted from the literature (known and candidate genes) and existing NGS datasets (public databases or through collaboration). The variants are evaluated in each of three evidential categories by using predefined criteria that will be manually curated: 1) "Population Genetic evidence" (e.g. population frequency, family segregation) and 2) "Functional Genetic evidence" (e.g. binding specificity, RNA expression) extracted from literature and data repositories and 3) "*in-silico* evidence" (e.g. PolyPhen2 for missense, ESEfinder for splice, RegulomeDB and GWAVA for non-coding variants, conservation measures). Each variant will then be placed in one of six ranks, based on the strength of evidence from the three categories [Table 1]. For each variant, summary data across the available datasets will be included rather than individual level data. Currently, data on all variants reported in all known PD genes in literature or identified in the WES at Miami UPDRCE have been imported in the database (~674k variants). All ranks and supporting data will be available to the user through the website (launch mid2015, online by ASHG meeting 2015). Additional summary data for each dataset and variant (e.g. QC, frequency) will be linked to the contributing lab's info or publication to facilitate collaborative efforts.

This database will allow users to quickly evaluate variants when interpreting sequence data, leading to a more focused follow-up in a research or clinical setting. The summary data format avoids the presence of individual identifiers, facilitating rapid availability of NGS data to the PD research community and thus potential collaborative efforts.

Table 1. Rankings for variants

	Rank 1	**Rank 2**	**Rank 3**
Genetic support	Highest	High	High
In-silico support	High	High or Intermediate	High or Intermediate
Functional support	Any	High	Intermediate or Low
	Rank 4	**Rank 5**	**Rank 6**
Genetic support	Intermediate	Low	Low
In-silico support	Any	High	Intermediate or Low

Functional support	Any	Intermediate or Low	Low

Dr. John T. Macdonald Foundation Department of Human Genetics
&
John P. Hussman Institute for Human Genomics

PRESENT THE

Human Genetics and Genomics Seminar Series

Karen Nuytemans, Ph.D.
Assistant Scientist

John P. Hussman Institute
for Human Genomics

Research Summary:

Through the use of next generation sequencing which allows for genome-wide rare variant discovery, the Parkinson Disease project focuses on identification of rare variants and their differential accumulation within a gene region or pathway in non-monogenic PD patients versus controls.

TOPIC:

"Insights into the complex nature of Parkinson Disease through next generation sequencing"

DATE:

Monday, December 1, 2014

TIME:

4:00p - 5:00p

LOCATION:

Biomedical Research Building (BRB)
John P. Hussman Institute
1501 NW 10th Avenue
3rd Floor Atrium

FOR MORE INFORMATION, PLEASE CONTACT
Dori McLean (dmclean@med.miami.edu)

UHealth UNIVERSITY OF MIAMI HEALTH SYSTEM
UNIVERSITY OF MIAMI MILLER SCHOOL of MEDICINE

From: Shepherd, Ruth [mailto:Shepherd.Ruth@mayo.edu]
Sent: Wednesday, March 11, 2015 10:44 AM
To: Mayo Clinic ListServ
Subject: Neurology Grand Rounds presents: Karen Nuytemans, PhD on March 16th, 12:15PM

Please see updated title and adjust your announcements accordingly. Thank you!

MAYO CLINIC

Neurology Grand Rounds

Kinne Auditorium

Monday March 16th, 2015 at 12:15 PM
Video-conferenced to Nemours 1 North & Mayo Clinic Health System in Waycross

Hosted by Dr. Zbigniew Wszolek

Special guest speaker Karen Nuytemans, PhD, presents *"Identifying Variants in Big Datasets and Assessing their Involvement in Parkinson's Disease"* at Neurology Grand Rounds
12:15 PM on March 16, 2015

Karen Nuytemans is an assistant scientist at the Miami Udall Center, John P. Institute of Human Genomics, University of Miami Miller School of Medicine. Throughout her training in Belgium (under Christine van Broeckhoven at the University of Antwerp) and in Miami Udall Center (under Dr. Vance at University of Miami), she has gained experience in characterizing Parkinson's Disease datasets for known and novel genes on both single base as copy number (CN) level. This experience led to the creation of PDmutDB, a database for variants

in the major PD genes based on literature. During her three year fellowship in Miami, she has extended her knowledge and acquired vast experience working with next generation sequencing (NGS) data and analyzing large datasets. At the Miami Udall Center, they are utilizing NGS to identify rare variant (accumulation) in PD cases versus controls. This expertise with rare variants has prompted them to extend upon the PDmutDB database to include NGS-identified variants and assess their impact in PD development.

Learning Objectives:

Define areas of new neuroscience knowledge and research

Understand Clinicopathologic (CPC) correlations of neurologic disease

Illuminate areas of practice-based improvement within the neurosciences based on advancing scientific research or Practice-based improvement projects.

This speaker, Karen Nuytemans, PhD, does not have a relevant financial relationship, and does not intend to discuss off label/investigative use of a commercial product or device.

College of Medicine, Mayo Clinic, is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

College of Medicine, Mayo Clinic, designates this live CME activity for a maximum of 1.0 *AMA PRA Category 1 Credits*[RM] . Physicians should claim only the credit commensurate with the extent of their participation in the activity.

This program is supported in part by an educational grant from the following companies in accordance with ACCME Standards: NONE

REFERENCES

1. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249.
2. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-1081.
3. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7(10):e46688.
4. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26(16):2069-2070.
5. Mort M, Sterne-Weiler T, Li B, et al. MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biol*. 2014;15(1):R19-2014-15-1-r19.
6. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003;31(1):3568-3571.
7. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67.
8. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004;11(2-3):377-394.
9. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997;4:311-323.
10. Rogozin IB, Milanesi L. Analysis of donor splice sites in different eukaryotic organisms. *J Mol Evol*. 1997;45(1):50-59.
11. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol*. 2011;18(10):1139-1146.
12. Griffiths-Jones S. miRBase: microRNA sequences and annotation. *Curr Protoc Bioinformatics*. 2010;Chapter 12:Unit 12.9.1-10.
13. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315.
14. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797.
15. Fu Y, Liu Z, Lou S, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol*. 2014;15(10):480.
16. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat Methods*. 2014;11(3):294-296.